Working towards human-level intelligence…

Dileep George

Vicarious
Intelligence:

Ability to model the world and to act on it
Intelligence:

Ability to **model** the world and to **act** on it
It is possible to act on the world without modeling it

Old brain
old brain

-600 million years

sponges, jellyfish, flatworm

-450

fish

-300

amphibians

-150

reptiles

birds

0

mammaliaforms

non-human primates

modern humans

earliest hominoids

new brain
<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dinosaurs appeared</td>
<td>Jan 1st</td>
</tr>
<tr>
<td>Dinosaurs became extinct</td>
<td>Sep 21st</td>
</tr>
<tr>
<td>Homo sapiens appeared</td>
<td>Dec 31st</td>
</tr>
</tbody>
</table>

Old brain was very successful. Dinosaurs had walnut-sized brains.
Old-brain style (model-free) learning

Stimulus → Trainable Function → Desired Response
Old-brain style (model-free) learning

- Evolutionary algorithms
- Back-propagation
- Reinforcement learning

Stimulus → Trainable Function → Desired Response
Old-brain style (model-free) learning

- Evolutionary algorithms
- Back-propagation
- Reinforcement learning

Stimulus → Trainable Function → Desired Response

Very large repertoire of training data
Old-brain style (model-free) learning

Stimulus → Trained Function → Predicted Response
Old-brain style (model-free) learning

Stimulus → Trained Function ↓ Look up table / Hash table → Predicted Response
Old-brain style (model-free) learning

- Trained Function
  - Look up table / Hash table
  - Predicted Response
Old-brain style (model-free) learning

Trained Function

Look up table / Hash table
Old-brain style (model-free) learning

Trained Function

Black box with limited capability for thinking, imagination, creativity or planning
Many modern methods are still like the old brain

- Extremely large number of training examples
- Inscrutable, black box classifiers
- Not generative. Models lack explanatory power
Deep Neural Networks are Easily Fooled:
High Confidence Predictions for Unrecognizable Images

<table>
<thead>
<tr>
<th>Anh Nguyen</th>
<th>Jason Yosinski</th>
<th>Jeff Clune</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Wyoming</td>
<td>Cornell University</td>
<td>University of Wyoming</td>
</tr>
<tr>
<td><a href="mailto:anguyen8@uwyo.edu">anguyen8@uwyo.edu</a></td>
<td><a href="mailto:yosinski@cs.cornell.edu">yosinski@cs.cornell.edu</a></td>
<td><a href="mailto:jeffclune@uwyo.edu">jeffclune@uwyo.edu</a></td>
</tr>
</tbody>
</table>

![Figure 1: Fooling images that are unrecognizable to humans](image-url)

Figure 1: Fooling images that are unrecognizable to humans.
starfish
Perhaps, the brain can tell us more about how to build truly intelligent machines.
Neocortex

Old brain
Assumptions  Learning efficiency  Generality
By discovering neocortex, nature stumbled upon the ‘magic middle’ of learning architectures.
This ‘magical architecture’ is:

General enough to solve multiple problems like vision, audition, somatosensory etc.

But specific enough to learn efficiently.

(This also implies that the same set of architectural assumptions work well for multiple domains)
Your brain is not very good at recognizing QR codes
Complex cells (Feature pooling)

Simple Cells (Feature detection)

Hubel & Wiesel
Neuroscience has had many more advances since 1961. Can’t we use that?
How?
How can we decipher information-processing principles from the brain?
Neocortex
Source of assumptions/constraints
Physics of World’s Data
To find correspondence with neocortex properties
Neocortex
Source of assumptions/constraints

Physics of World’s Data
To find correspondence with neocortex properties

Computational Framework
Understand why neocortex does what it does to design algorithms
Observed hierarchy in the cortex

Hierarchical structure of data

Efficiency and re-use.
Machine learning, statistics
Brain corporation
New-brain research questions

• Hierarchy + Feedback + Temporal Learning & Inference
• Scene understanding
• Sensori-motor integration
• Concept learning
• Language understanding
Vision System

Conceptual System and Language

Motor System
How we solved CAPTCHAs
We stuck to new-brain methods:

- Small amounts of training data
  - Trained on clean examples
  - Trained only on positive examples
- Unsupervised training
- ....
Few training examples

**Interpolation** of millions of training images

**VS**

**Extrapolation** from few examples
Solves All Variations

Total training examples < 1000 !
Our algorithm produces very detailed segmentation
even when the contours are occluded
a
ICDAR

short

hungry

Campus

animal

hot

BEER
Scene parsing
Systems with imagination..
What is this picture?
It is the picture of a bear climbing a tree. Can you see the bear?
Imaginations from our system!
Why is imagination important for AGI?
Vision System

Motor System

Conceptual System and Language
Wind Tunnel

Kite

Glider
Major breakthroughs in A.I will need some wind-tunnel style research
When?
Will we solve the fundamental research problems in $N$ years?

<table>
<thead>
<tr>
<th>$N \leq 5$</th>
<th>No way</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5 &lt; N \leq 10$</td>
<td>Small possibility</td>
</tr>
<tr>
<td>$10 &lt; N \leq 20$</td>
<td>$&gt; 50%$</td>
</tr>
</tbody>
</table>
Increase in brain size during human evolution

- Homo sapiens sapiens
- Homo sapiens Neandertalensis
- Homo erectus
- Homo habilis
- Australopithecus africanus
- Australopithecus (Singe Anthropoide)

Millions of years ago:
- 4
- 3
- 2
- 1
- 0

Brain size:
- 500 cc
- 1000 cc

Us!
About intelligence explosion…
Self-limiting forces on A.I that will prevent an uncontrolled growth
• Data limit
• Intervention limit
• ...
• **Data limit**
  
  • Super-human A.I will encounter limits of available data
  
  • Limit on recursive improvement
    
    • Going over the same data again and again does not increase the information in the data. (Data Processing Inequality)

  • Natural dynamics of data generation
Suppose super-human A.I was created before we had any knowledge about how gravity worked and how earth moved around the sun.

How long will it take for the A.I to discover that it takes the earth 365 days to go around the sun?

The dynamics of the world imposes limits on how fast data can be acquired.